UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

March 9, 2001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES #V-5

MEMORANDUM FOR Dennis Stoudt
Assistant Division Chief, Processing and Support
Decennial Systems and Contracts Management Office

From: Donna Kostanich DK
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Michael Starsinic MS and Jae Kwang Kim JK
Long Form and Variance Estimation Staff

Subject: Accuracy and Coverage Evaluation: Computer Specifications for Variance Estimation for Census 2000 - Revision

The attachment contains the *final* computer programming specifications for calculating the A.C.E. post-stratum level variances and computing the covariance matrix. Please direct any questions about these specifications to Michael Starsinic or Jae Kwang Kim at 457-1638.

These specifications assume:

- All correct enumeration, residence, and match probabilities have been assigned,
- All phases of the P and E-Sample missing data process have been completed, and
- Final P and E sample weights have been calculated and post-strata DSE estimates have been computed.

Additional specifications will be provided for producing generalized variance estimates.

Note: This specification supersedes DSSD Census 2000 Procedures and Operations Memorandum Series #V-1, "Computer Specifications for Variance Estimation for Census 2000".

Attachment

COMPUTER SPECIFICATIONS FOR VARIANCE ESTIMATION FOR CENSUS 2000

I. INTRODUCTION

A. Variance Components

Replication methods will be used to estimate the variance due to A.C.E. sampling and estimation. These variance estimates will reflect three components of the variation of the census estimates:

1. Variance due to the multi-phase sampling of block clusters for the A.C.E., specifically the Initial Listing sample, A.C.E. Reduction, and Small Block Subsampling.

2. Variance due to sampling for the Targeted Extended Search (TES).

3. Variance from estimating the missing data in A.C.E.

B. Contents

This specification describes the variance estimation process for Census 2000. It contains two major parts. These are:

- Overview of A.C.E. Sampling & Estimation, and TES
- Computation of the A.C.E. Variance

II. OVERVIEW OF A.C.E. ESTIMATION, A.C.E. SAMPLING & TES

A. A.C.E. Sampling

Block clusters were initially stratified by size of cluster and whether the cluster was on an American Indian Reservation (AIR): 0-2 housing units ("small" block clusters), 3-79 housing units and not on an AIR ("medium" block clusters), 80 or more housing units and not on an AIR ("large" block clusters), and 3 or more housing units on an AIR ("AIR" block clusters; not all states have this stratum). Systematic samples were selected independently in each state.[1]

Medium and large block clusters were subsampled to lower the total number of housing units in the A.C.E. sample to approximately 300,000. They were stratified based on minority/non-minority status crossed by three consistency codes. AIR block clusters and

[1] DSSD Census 2000 Procedures and Operations Memorandum Series R-3, "Accuracy and Coverage Evaluation (A.C.E.) Survey: Block Cluster Sample Selection Specification

Puerto Rico block clusters were not subsampled in this operation.[2] Small blocks were also subsampled in a separate operation from the A.C.E. reduction sample above.[3] No block clusters were eligible to be subsampled in both A..C.E. Reduction and Small Block Cluster Subsampling. Large block clusters were further subsampled. They were divided into segments, and these segments were subsampled to reduce enumerator workloads.[4]

B. A.C.E. ESTIMATION

Estimation of the total population of the United States is achieved by summing dual system estimates (DSE's) of mutually exclusive and exhaustive post-strata. The DSE equation remains the same regardless of the definition of the post-strata. The form of the DSE is:[5]

$$D\hat{S}E = (C^* - \Pi^*)\left(\frac{CE}{N_e}\right)\left(\frac{N_n + N_i}{M_n + \left(\frac{M_o}{N_o}\right)N_i}\right)$$

where : C^* = unweighted census count excluding late adds and deletes
Π^* = count of not-data-defined and wholly imputed persons
CE = estimated number of A.C.E. E-Sample correct enumerations
N_e = estimated number of A.C.E. E-Sample persons
N_n = estimated number of A.C.E. P-Sample nonmovers
N_i = estimated number of A.C.E. P-Sample inmovers
N_o = estimated number of A.C.E. P-Sample outmovers
M_n = estimated number of A.C.E. P-Sample nonmover matches
M_o = estimated number of A.C.E. P-Sample outmover matches

The DSE is computed separately for each post-stratum, i. The national DSE estimate is computed as:

$$D\hat{S}E_{US} = \sum_i D\hat{S}E_i$$

[2] DSSD Census 2000 Procedures and Operations Memorandum Series R-29, "Accuracy and Coverage Evaluation Survey: Reduction Specification"

[3] DSSD Census 2000 Procedures and Operations Memorandum Series R-24, "Accuracy and Coverage Evaluation Survey: Small Block Cluster Subsampling"

[4] DSSD Census 2000 Procedures and Operations Memorandum Series R-27, "Accuracy and Coverage Evaluation: Large Block Cluster Subsampling Specifications"

[5] DSSD Census 2000 Procedures and Operations Memorandum Series Q-20, "A.C.E. Dual System Estimation"

The Coverage Correction Factor (CCF) is:

$$\hat{CCF} = \frac{\hat{DSE}}{C}$$

where C is the Census count including possible late adds and deletes, and would be equal to C^* if there were none.

Puerto Rico, while handled separately, follows the same basic estimation process.

Missing data must be imputed for items needed for the estimation process. Item nonresponse is imputed through either hot deck or ratio methods. Unresolved residence and match status in the P-sample, and unresolved correct enumeration status for the E-sample are also imputed using ratio methods. Households selected but not interviewed in the A.C.E. are compensated for by using a noninterview weight adjustment.[6] Variance due to imputation for item nonresponse is not included in the variance estimate, but the variance due to unresolved match, residence and correct enumeration probability and noninterview adjustment estimation is.

C. TES SAMPLING

For the A.C.E., 20 percent of clusters will have their surrounding blocks searched for additional matches and correct enumerations of persons who may have been misplaced geographically, as opposed to 100 percent of clusters in the 1990 Post-Enumeration Survey. Approximately 10 percent will be selected with certainty, including relist clusters and clusters with high numbers of weighted or unweighted nonmatches and geocoding errors. The remaining 10 percent will be taken from a systematic sample of the remaining clusters that contain at least one "interesting housing unit" (a unit coded in Initial Housing Unit Matching as a geocoding error, or a housing unit on the Independent List which did not match to a census unit).[7]

III. COMPUTATION OF THE A.C.E. VARIANCES

A. Overview

For Census 2000, a standard stratified jackknife such as was used in the Census 2000 Dress Rehearsal cannot be used because it cannot properly take into account the variance due to the multi-phase nature of the sampling. A multi-phase sample, as opposed to a multi-stage sample, uses information from the first sample's results to draw the second

[6] Ikeda, Michael. Internal memorandum, "Overview of Proposed Missing Data Procedures for the Census 2000 Accuracy and Coverage Evaluation Sample"

[7] DSSD Census 2000 Procedures and Operations Memorandum Series R-20, "Accuracy and Coverage Evaluation Survey- Identification and Sampling of Block Clusters for Targeted Extended Search"

sample (e.g. defining strata). A modification of the Rao-Shao jackknife variance estimator for a reweighted expansion estimator developed by Jae Kim will be used instead.[8]

The E and P-Sample A.C.E. person records are ordered in sampling strata x block cluster order. In the jackknife, a block cluster is deleted from one replicate, while the remaining block clusters in the same stratum are reweighted.

For each replicate, the following steps of estimation should be performed and implemented:

1. Recalculation of the missing data adjustments for missing P-sample match status, E-sample correctness of enumeration, and residence probability

2. Calculation of the dual-system estimate

After the jackknifing is finished, the final step is to create a variance-covariance matrix based on the collapsed post-strata.

B. Input Files

1. Post-Stratum Summary File

   An output of DSE estimation, this file contains the collapsed post-strata, population & II (wholly imputed person) counts for each cluster.

2. Sample Design File

   This file is needed to assign collapsed sampling strata to clusters, and get various other cluster-level information.

3. Missing Data Files

   Output files produced in P & E-Sample Missing Data Processing and modified during DSE estimation.

C. Output Files

1. Post-Stratum Group Detailed Variance Files

   These files, one for each of the 64 post-stratum groups (PSG=int(POSTSTR/10)), give the variance and associated results for all post-strata and post-strata groups.

[8] Kim, Jae Kwang. Internal memorandum, "Replication Variance Estimation for Multi-Phase Stratified Sampling"

2. National Variance File

This file contains the national-level variance and other national-level information.

3. DSE Variance-Covariance Matrix & CCF Variance-Covariance Matrix

These files contain a K x K variance-covariance matrix (where K is the number of collapsed post-strata). The CCF file will be crucial in computing generalized variances, which will be described in a forthcoming specification.

4. Replicate Weight File

This file contains the information needed to create the replicate weights (RW's) for each replicate for each cluster in sample. It does not contain all replicates (29,136 x 11,303 = 329,324,208) because of size constraints.

D. Creating the A.C.E. Replicate File

1. Create the Collapsed Post-Strata File

Extract the following variables from the Post-Collapsed Post-Stratum Summary File, an output of DSE estimation[9]:

Variable Name	Variable Description	Location in File
CPOSTSTR	Collapsed Post-Stratum Code	005-007
C*	HCEF* Census Count	009-018
II*	HCEF* Insufficient Information Persons	020-029
C	HCEF Census Count	246-255

Note that the first line of the file will contain national totals; the second line will contain post-stratum 001.

2. Create the Final Sampling Stratum Variable & Number the Clusters

From the Sample Design File, extract the block cluster records of all clusters in the A.C.E. sample before A.C.E. Reduction and Small Block Subsampling including those that were sampled out, i.e. those with BC2=1.

For each desired block cluster record, extract the following variables:

[9] DSSD Census 2000 Procedures and Operations Memorandum Series Q-?, "Accuracy and Coverage Evaluation Survey: Computer Specifications for Person Dual System Estimation Output Files"

Variable Name	Variable Description	Location in File
STATE	FIPS State Code	3-4
CLUST	A.C.E. Block Cluster Number	21-25
DIGIT	A.C.E. Block Cluster Check Digit	26
SS	Sampling Stratum: 1 = Small 2 = Medium 3 = Large 4 = Medium or Large on AIR	55
BC2	Second step listing sample selection indicator, 0 = not selected, 1 = selected	127
ARS	A.C.E. Reduction Stratum	190-191
SBCSS	Small Block Cluster Sampling Stratum	306-307
SB	Small Block Subsampling Indicator, 0 = not selected, 1 = selected	308
TESELECT	TES Selection Code: R, U, W - In Certainty Sample S - In Systematic Sample N - Not Sampled O - L/E Clusters I - Ineligible	694

SB is a variable which can identify the "surviving" clusters after Small Block Cluster Subsampling. All clusters that were selected in this sample also must have been selected in each previous step.

Create a new variable, Final Sampling Stratum (FSS). If SS = 1, then:

$$FSS = SBCSS + 10000*SS + 100000*STATE$$

If SS=2:

$$FSS = 10 + 100*ARS + 10000*SS + 100000*STATE$$

If SS=3:

$$FSS = 11 + 100*ARS + 10000*SS + 100000*STATE$$

If SS=4:

$$FSS = SBCSS + 100*ARS + 10000*SS + 100000*STATE$$

Also, let

ISS = SS + 10*STATE

Sort the clusters by FSS. Create the following new variables:

NBEF = consecutive numbering of all clusters from 1 to J (j-indexed), the total number of clusters before A.C.E. Reduction and Small Block Cluster Subsampling (29,136).

NAFT = consecutive numbering of all clusters with SB=1 (those that "survived" all levels of sampling) from 1 to I (i-indexed) , the total number of clusters after A.C.E. Reduction and Small Block Cluster Subsampling (11,303). This should be left blank for clusters with SB≠1.

Let the notation FSS(i) indicate the FSS of cluster i (similarly for ISS(i), FSS(j), and ISS(j)). Let the notation j(i) indicate the value of NBEF for the cluster with NAFT=i.

3. Collapse Final Sampling Strata

Some FSS's contain only one block cluster. In this situation the sampling stratum needs to be collapsed into another stratum. The collapsing has been predetermined. Follow the collapsing scheme in this table:

Original, Problem FSS	**Collapse Into This FSS**	**Original, Problem FSS**	**Collapse Into This FSS**
110002	2810002	3310002	5010002
210002	5010002	3410002	4210002
230411	5630411	3530211	430211
530411	2230411	3530411	430411
1010002	4410002	3830111	3030111
1130411	1030411	3830411	4630411
1910003	3110003	4010002	2010002
2110003	1310003	4510003	3710003
2330211	2530211	4630111	3030111
2410002	5110002	4930111	830111
2510002	2310002	5110003	3710003
2710002	3810002	5430111	2430111
3130111	2030111	5510003	1710003
3130411	1930411		

4. Create the Cluster-Level A.C.E. Replicate Weights

Replicate weights are the means by which the jackknife is implemented. One set of replicate weights will be needed. For cluster i and replicate j:

$$RW_i^{(j)} = \begin{cases} 0 & \text{if } j(i)=j \\ \dfrac{r_{CFSS,i}}{r_{CFSS,i}-1}\ \dfrac{n_{CFSS,i}-1}{n_{CFSS,i}}\ \dfrac{n_{ISS,i}}{n_{ISS,i}-1} & \text{if } j(i)\neq j,\ CFSS(i)=CFSS(j),\ SB(j)=1 \\ \dfrac{n_{CFSS,i}-1}{n_{CFSS,i}}\ \dfrac{n_{ISS,i}}{n_{ISS,i}-1} & \text{if } j(i)\neq j,\ CFSS(i)=CFSS(j),\ SB(j)\neq 1 \\ \dfrac{n_{ISS,i}}{n_{ISS,i}-1} & \text{if } j(i)\neq j,\ ISS(i)=ISS(j),\ CFSS(i)\neq CFSS(j) \\ 1 & \text{if } j(i)\neq j,\ ISS(i)\neq ISS(j),\ CFSS(i)\neq CFSS(j) \end{cases}$$

where:

$n_{ISS,i}$ = the number of clusters in the same original (*not* collapsed) ISS as cluster i

$n_{CFSS,i}$ = the number of clusters in the same Collapsed FSS as cluster i

$r_{CFSS,i}$ = the number of clusters in the same Collapsed FSS as cluster i with SB=1

If $n_{ISS,i}$, $n_{CFSS,i}$, or $r_{CFSS,i}$ is equal to one, then set the ratio involving that value to one.

There will be a total of J+1 replicates, with the extra being the zeroth replicate, which represents the full A.C.E. sample and for which all replicate weights≡1. Usually in jackknifing, the number of observations over which the estimates are calculated equals the number of replicates, but here we only need to use the I clusters which are still in sample after the second phase of sampling (i.e. those with SB=1).

All persons in a cluster receive the same weight.

5. Create the E- and P-Sample Replicate Files

   a. Extract the following variables from the E-Sample Person Missing Data Output File:

E-Sample Missing Data Output File Extract

Variable Name	Variable Description	Location in File
CLUSTER	Cluster Number and Check Digit	17-22
RACE	63-Category Race Code	83-84
PSPAN	Hispanic Code	86
NUMEDUP	Number of Duplicates with E-Sample Persons	89

NUMDUP	Number of Duplicates with Non-E-Sample Persons	90-91
FINMAT	Final Match Code	101-102
TESPER	TES Person Indicator	104
EWGHT	E-Sample Trimmed Weight	110-123
TESWGT	TES Sampling Weight	130-136
TENURE2	Recoded/Imputed Tenure	159
AGE2	Recoded/Imputed Age	160
RELATE2	Recoded Relationship	163
AMTIMP	Flag for Item Imputation	169
BFUGP	E-Sample Before Follow-up Match Code Group	170-171
CEPROBI	Initial Probability of Correct Enumeration	177-186
CEPROBF	Final Probability of Correct Enumeration	187-197
POSTSTR	Post-stratum Code	311-313
CPOSTSTR	Collapsed Post-stratum Code	314-316

Extract the following variables from the P-Sample Person Missing Data Output File:

P-Sample Missing Data Output File Extract

Variable Name	Variable Description	Location in File
CLUSTER	Cluster Number and Check Digit	17-22
RACE	63-Category Race Code	100-101
MOVERPER	Person Mover Flag	107
RSC	Computer Residence Status Code	109
FINMAT	Final Match Code	121-122
TESPER	TES Person Indicator	125
ADDCDE	Address Code	128
TESWGT	TES Sampling Weight	129-135
TENURE2	Recoded/Imputed Tenure	146
AGE2	Recoded/Imputed Age	147

HISP2	Recoded/Imputed Hispanic Origin	150
RELATE2	Recoded Relationship	152
AMTIMP	Flag for Item Imputation	158
BFUGP	P-Sample Before Follow-up Match Code Group	159-160
RPROB	Probability of Residence	166-175
MPROB	Match Probability	176-186
PWGHT	Initial P-Sample Trimmed Weight	187-200
NIWGTO	P-Sample Noninterview Adjusted Weight based on Census Day	201-215
NIWGTI	P-Sample Noninterview Adjusted Weight based on Interview Day	216-230
TESFINWT	Final weight used in estimation	231-245
POSTSTR	Post-stratum Code	345-347
CPOSTSTR	Collapsed Post-stratum Code	348-350

b. Append the variables from steps 2 and 3 (merge by A.C.E. Cluster Number and Check Digit) to both files.

c. Append n_{ISS}, n_{FSS}, r_{FSS}, and the J+1 replicate weights (or equivalents) from step 4 to both the E- and P-Sample files.

E. Compute the A.C.E. Variances

1. Replicate Missing Data Imputation

a. Overview

For each replicate, adjust the imputed values of CEPROBF on the E-Sample File, and MPROB and RPROB on the P-Sample file based on the replicate weights. The general form for the imputation of the probability for an unresolved person in imputation cell j is:

$$Pr_j^* = \frac{\sum_{\text{resolved units} \in j} w_i \, w_i^* \, Pr_i}{\sum_{\text{resolved units} \in j} w_i \, w_i^*}$$

where

w_i = person-level weight, incorporating all levels of sampling and weight trimming, but not including noninterview adjustment

$$w_i^* = \begin{cases} \text{conditional TES weight, the inverse of the probability of selection} \\ \quad \text{in the TES sample (TESWGT on the P- and E-Sample Missing} \\ \quad \text{Data files) if the person is a TES person} \\ \text{1 if the person is NOT a TES person} \end{cases}$$

$$Pr_j = \begin{cases} \text{1 if a person is a \{match/resident/correct enumeration\}} \\ \text{0 if a person is NOT a \{match/resident/correct enumeration\}} \end{cases}$$

If the denominator of Pr_j^* is zero in a specific imputation cell, then set Pr_j^* to 1.0 in that cell. If there are no unresolved units in an imputation cell, skip it.

b. Non-Interview Adjustment (P-Sample File Only)

The Non-Interview Adjustment factor does not need to be recalculated. The calculation of the adjustment is done within block cluster, and since our replicate weights are defined at the block cluster-level, the replicate weights cancel out. Use the Noninterview Adjusted Weights as they are on the input file.

c. Residence Probability (P-Sample File Only)

For each replicate j and imputation cell m, calculate the imputed residence probability of each imputation cell as

$$RPROB_m^{*(j)} = \frac{\sum\limits_{\text{resolved persons} \in m} RW_p^{(j)}\, PWGHT_p\, TESWGT_p^{(j)}\, RPROB_p}{\sum\limits_{\text{resolved persons} \in m} RW_p^{(j)}\, PWGHT_p\, TESWGT_p^{(j)}}, \text{ where}$$

$$TESWGT_p^{(j)} = \begin{cases} \left(\dfrac{\sum\limits_{i \in TESELECT=S,N}^{I} RW_i^{(j)}}{\sum\limits_{i \in TESELECT=S}^{I} RW_i^{(j)}} \right) \text{ if } TESELECT_p = S \text{ and } TESPER_p = 1 \\ TESWGT_p \text{ otherwise} \end{cases}$$

Resolved units have FINMAT = M, MR, NP, NC, NR, FP, NL, NN, DP, MN, or GP.

The imputation cells for estimation of P-Sample residence probability are defined below.

BFUGP	Owner				Non-Owner			
	NH White		Others		NH White		Others	
	V3a	V3b	V3a	V3b	V3a	V3b	V3a	V3b
1	1		2		3		4	
2	5		6		7		8	
3	9	10	11	12	13	14	15	16
4	17		18		19		20	
5	21		22		23		24	
6	25		26		27		28	
7	29*		30*		31*		32*	
8	33		34		35		36	

Match code groups are defined above. Non-Hispanic White is defined by HISP2=1 (Non-Hispanic) and RACE=32 (White, not in combination with other races). Tenure is defined by TENURE2: TENURE2=1 is Owner, TENURE2=2 is Non-Owner. V3 is a variable defined for BFUGP=3 only. V3a includes persons in group 3 with AGE2=2 (18-29) and RELAT2=2 (Child of Reference Person). V3b includes all other persons with BFUGP=3.

Residence probabilities for cells 1-28 and 33-36 are calculated using only resolved persons within that cell. However, cells 29-32 are computed using resolved persons with the appropriate race/origin/tenure characteristics in BFUGP's 1 through 5. For example, the imputed probability for cell 29 is calculated from resolved persons in cells 1, 5, 9, 10, 17 and 21.

d. Match Probability (P-Sample File Only)

For each replicate j and imputation cell m, calculate the imputed match probability of each imputation cell as

$$\mathrm{MPROB}_m^{*(j)} = \frac{\sum_{\text{resolved residents} \in m} \mathrm{RW}_p^{(j)}\,\mathrm{PWGHT}_p\,\mathrm{TESWGT}_p^{(j)}\,\mathrm{MPROB}_p}{\sum_{\text{resolved residents} \in m} \mathrm{RW}_p^{(j)}\,\mathrm{PWGHT}_p\,\mathrm{TESWGT}_p^{(j)}}, \text{ where}$$

$$\mathrm{TESWGT}_p^{(j)} = \begin{cases} \left(\dfrac{\sum_{i \in \mathrm{TESELECT}=S,N}^{I} \mathrm{RW}_i^{(j)}}{\sum_{i \in \mathrm{TESELECT}=S}^{I} \mathrm{RW}_i^{(j)}} \right) & \text{if } \mathrm{TESELECT}_p = S \text{ and } \mathrm{TESPER}_p = 1 \\ \mathrm{TESWGT}_p \text{ otherwise} & \end{cases}$$

Resolved residents have FINMAT = M, MR, MU, NP, NC, NR, or NU.

The imputation cells for estimation of P-Sample match probability are defined below.

Mover Status	HU match		HU Nonmatch or Conflicting HH	
	0 imputes	**1+ imputes**	**0 imputes**	**1+ imputes**
Non-mover	1	2	3	4
Out-mover	5	6	7	

Housing-Unit Address Matchcode is determined by ADDCDE. ADDCDE=1 indicates a HU match; ADDCDE=2 or 4 indicates a nonmatch or conflicting HH. Mover status is determined by MOVERPER. MOVERPER=1 indicates non-mover; MOVERPER=3 indicates out-mover. The variable AMTIMP (0 imputes or 1+ imputes in the table) denotes how many of the following characteristics were imputed for the P-Sample person: age, sex, race, hispanic origin, and tenure. AMTIMP=0 indicates no imputes, AMTIMP$\geq$1 indicates at least one impute.

e. Correct Enumeration Probability (E-Sample File Only)

Imputation for correct enumeration probability is a two-step process. First, for each replicate j and imputation cell m, calculate the imputed initial correct enumeration probability of each imputation cell as

$$\mathrm{CEPROBI}_m^{*(j)} = \frac{\displaystyle\sum_{\text{resolved persons} \in m} \mathrm{RW}_p^{(j)}\,\mathrm{PWGHT}_p\,\mathrm{TESWGT}_p^{(j)}\,\mathrm{CEPROBI}_p}{\displaystyle\sum_{\text{resolved persons} \in m} \mathrm{RW}_p^{(j)}\,\mathrm{PWGHT}_p\,\mathrm{TESWGT}_p^{(j)}},$$

$$\text{where } \mathrm{TESWGT}_p^{(j)} = \begin{cases} \left(\dfrac{\displaystyle\sum_{i \in \mathrm{TESELECT}=S,N}^{I} \mathrm{RW}_i^{(j)}}{\displaystyle\sum_{i \in \mathrm{TESELECT}=S}^{I} \mathrm{RW}_i^{(j)}} \right) & \text{if } \mathrm{TESELECT}_p\text{=S and } \mathrm{TESPER}_p=1 \\ \mathrm{TESWGT}_p \text{ otherwise} & \end{cases}$$

Here, resolved is defined as FINMAT = M, CE, MR, GE, EE, FE, DE, MN, or KE.

The imputation cells for estimation of E-Sample correct enumeration probability are defined below.

BFUGP	0 Imputes	1+ Imputes
1	1	2

2	3	4
5	5	6
6	7	8
10	23	24
11	25	26
12	27	28

BFUGP	**0 Imputes**		**1+ Imputes**	
	V3a	**V3b**	**V3a**	**V3b**
3	9	10	11	12

BFUGP	**0 Imputes**		**1+ Imputes**
	NH White	**Others**	**All Resolved Persons**
4	13	14	15
7	16	17	18
8	19	20	21

BFUGP	**All Resolved Persons**
9	22*

Non-Hispanic White is defined by PSPAN = 1 (Non-Hispanic) and RACE = 32 (White). 0 imputes is defined by AMTIMP = 0 (no variables imputed in either E-Sample or HCEF). V3a includes persons in match code group 3 with AGE2 = 2 (18-29) and RELAT2 = 2 (Child of Reference Person). V3b includes all other persons in match code group 3.

For unresolved persons with BFUGP=9 (cell 22), do not use the formula above. Set $\mathrm{CEPROBI}_{22}^{*(j)}=0$.

Next, for each unresolved *individual*,

$$\mathrm{CEPROBF}_p^{*(j)} = \mathrm{CEPROBI}_p^{*(j)} \times \frac{\mathrm{NUMEDUP}_p + 1}{\mathrm{NUMEDUP}_p + \mathrm{NUMDUP}_p + 1}$$

f. For each individual p in the P-Sample, and for the jth replicate:

$$\mathrm{RPROB}_p^{(j)} = \begin{cases} \mathrm{RPROB}_p & \text{if resolved} \\ \mathrm{RPROB}_p^{*\,(j)} & \text{if unresolved (MU, NU, P, KI, KP)} \end{cases}$$

$$\mathrm{MPROB}_p^{(j)} = \begin{cases} \mathrm{MPROB}_p & \text{if resolved} \\ \mathrm{MPROB}_p^{*\,(j)} & \text{if unresolved (P, KI, KP)} \end{cases}$$

For each individual p in the E-Sample, and for the jth replicate:

$$\mathrm{CEPROBF}_p^{(j)} = \begin{cases} \mathrm{CEPROBF}_p & \text{if resolved} \\ \mathrm{CEPROBF}_p^{*\,(j)} & \text{if unresolved (UE, MU, P, GU)} \end{cases}$$

2. Replicate DSE Estimation and Compute Variances

The basic formula for DSE estimation is:

$$\mathrm{DSE} = (C^* - \mathrm{II}^*)\left(\frac{CE}{N_e}\right)\left(\frac{N_n + N_i}{M_n + \left(\frac{M_o}{N_o}\right) N_i}\right)$$

where:

$$N_e = \sum_{p \in \text{E-Sample}} \mathrm{EWGHT}_p$$

$$CE = \sum_{p \in \text{E-Sample}} \mathrm{CEPROBF}_p \times \mathrm{EWGHT}_p$$

$$N_n = \sum_{p \in \text{Nonmovers}} \mathrm{RPROB}_p \times \mathrm{NIWGTO}_p$$

$$N_o = \sum_{p \in \text{Outmovers}} \mathrm{RPROB}_p \times \mathrm{NIWGTO}_p$$

$$N_i = \sum_{p \in \text{Inmovers}} \mathrm{NIWGTI}_p$$

$$M_n = \sum_{p \in \text{Nonmovers}} \mathrm{MPROB}_p \times \mathrm{RPROB}_p \times \mathrm{NIWGTO}_p$$

$$M_o = \sum_{p \in \text{Outmovers}} \mathrm{MPROB}_p \times \mathrm{RPROB}_p \times \mathrm{NIWGTO}_p$$

If the denominator of any ratio is equal to zero, set that ratio equal to 1.

a. To estimate the national DSE variance, first compute:

$$\mathrm{Term}_{h,k}^{(j)} = \sum_{i=1}^{I}\left(\mathrm{RW}_i^{(j)}\sum_{p\in h} w_{ipk}\, x_{ip}\right) + \sum_{i=1}^{I}\left(\mathrm{RW}_i^{(j)}\sum_{p\in h} w_{ipk}\, y_{ip}\right)$$

$$+ \left(\frac{\sum\limits_{i\in \mathrm{TESELECT=S,N}}^{I} \mathrm{RW}_i^{(j)}}{\sum\limits_{i\in \mathrm{TESELECT=S}}^{I} \mathrm{RW}_i^{(j)}}\right)\sum_{i=1}^{I}\left(\mathrm{RW}_i^{(j)}\sum_{p\in h} w_{ipk}\, z_{ip}\right)$$

$$= \sum_{i=1}^{I}\left\{\mathrm{RW}_i^{(j)}\sum_{p\in h} w_{ipk}\left[x_{ip} + y_{ip} + \left(\frac{\sum\limits_{i\in \mathrm{TESELECT=S,N}}^{I} \mathrm{RW}_i^{(j)}}{\sum\limits_{i\in \mathrm{TESELECT=S}}^{I} \mathrm{RW}_i^{(j)}}\right) z_{ip}\right]\right\}$$

If the denominator of the third component above is zero, set the ratio equal to one.

A "Term" is any of the seven individual numerators and denominators, excluding C^* and Π^*, making up the DSE formula, and:

h = post-stratum designator
k = term designator
j = "j-indexed" replicate designator (NBEF)
i = "i-indexed" cluster designator (NAFT)
I = the number of clusters after subsampling is completed
p = person designator
w_{ipk} = final weight, excluding the TES weight, where "weight" varies by term k:

k	Term	"Final Weight", w_{ipk}	Restriction
1	$N_e^{(j)}$	EWGHT_p	
2	$CE^{(j)}$	$\mathrm{CEPROBF}_p^{(j)}$ x EWGHT_p	
3	$N_n^{(j)}$	$\mathrm{RPROB}_p^{(j)}$ x NIWGTO_p	MOVERPER=1 (Nonmovers only)
4	$N_o^{(j)}$	$\mathrm{RPROB}_p^{(j)}$ x NIWGTO_p	MOVERPER=3 (Outmovers only)
5	$N_i^{(j)}$	NIWGTI_p	MOVERPER=2 & RSC≠R (Inmovers only)
6	$M_n^{(j)}$	$\mathrm{MPROB}_p^{(j)}$ x $\mathrm{RPROB}_p^{(j)}$ x NIWGTO_p	MOVERPER=1 (Nonmovers only)
7	$M_o^{(j)}$	$\mathrm{MPROB}_p^{(j)}$ x $\mathrm{RPROB}_p^{(j)}$ x NIWGTO_p	MOVERPER=3 (Outmovers only)

x_{ip} = 1 if the person is NOT a TES person (TESPER=0), 0 otherwise
y_{ip} = 1 if the person IS a TES person AND is from a cluster sampled with certainty in TES (TESPER=1 and TESELECT=O, R, U, or W), 0 otherwise
z_{ip} = 1 if the person IS a TES person AND is from a non-certainty sampled cluster (TESPER=1 and TESELECT=S), 0 otherwise

b. Using the P-sample file, tally the number of outmovers in each post-stratum (MOVERPER=3, RPROB>0, and TESFINWT>0). If there are 10 or more outmovers in post-stratum h, then for all replicates of that post-stratum use the PES-C formula for the DSE:

$$D\hat{S}E_h^{(j)} = (C_h^* - II_h^*)\left(\frac{CE_h^{(j)}}{N_{e,h}^{(j)}}\right)\left(\frac{N_{n,h}^{(j)} + N_{i,h}^{(j)}}{M_{n,h}^{(j)} + \left(\frac{M_{o,h}^{(j)}}{N_{o,h}^{(j)}}\right) N_{i,h}^{(j)}}\right)$$

If there are fewer than 10 outmovers in post-stratum h, then for all replicates of that post-stratum use the PES-A formula for the DSE

$$D\hat{S}E_h^{(j)} = (C_h^* - II_h^*)\left(\frac{CE_h^{(j)}}{N_{e,h}^{(j)}}\right)\left(\frac{N_{n,h}^{(j)} + N_{o,h}^{(j)}}{M_{n,h}^{(j)} + M_{o,h}^{(j)}}\right)$$

Note that in this situation, it is not necessary to compute term 5, $N^{(j)}_{i,h}$.

c. The DSE variance estimate for post-stratum h is:

$$\mathrm{Var}(D\hat{S}E_h) = \sum_{j=1}^{J} \frac{n_{ISS,j}-1}{n_{ISS,j}} (D\hat{S}E_h^{(j)} - D\hat{S}E_h^{(0)})^2; \quad \text{if } n_{ISSj} = 1, \text{ set } \frac{n_{ISS,j}-1}{n_{ISS,j}} = 1$$

The variance of the CCF for post-stratum h is:

$$\mathrm{Var}(C\hat{C}F_h) = \frac{\mathrm{Var}(D\hat{S}E_h)}{C_h^2}$$

d. Compute the post-stratum DSE & CCF variance-covariance matrices. The DSE covariance between post-strata h and h' is:

$$\mathrm{Cov}(D\hat{S}E_h, D\hat{S}E_{h'}) = \sum_{j=1}^{J} \frac{n_{ISS,j}-1}{n_{ISS,j}} (D\hat{S}E_h^{(j)} - D\hat{S}E_h^{(0)})(D\hat{S}E_{h'}^{(j)} - D\hat{S}E_{h'}^{(0)})$$

If $n_{ISS,j} = 1$, then set $\frac{n_{ISS,j} - 1}{n_{ISS,j}} = 1$.

The CCF covariance between post-strata h and h' is:

$$\mathrm{Cov}(C\hat{C}F_h, C\hat{C}F_{h'}) = \frac{\mathrm{Cov}(D\hat{S}E_h, D\hat{S}E_{h'})}{C_h C_{h'}}$$

e. Compute the national variance estimate.

$$\mathrm{Var}(\hat{DSE}_{US}) = \sum_{h=1}^{448} \sum_{h'=1}^{448} \mathrm{Cov}(\hat{DSE}_h, \hat{DSE}_{h'})$$

where,

$$\mathrm{Cov}(\hat{DSE}_h, \hat{DSE}_h) = \mathrm{Var}(\hat{DSE}_h)$$

F. Creating Variance Output Files

1. Post-Stratum Group Detailed Variance File

Sixty-four of these tables will be produced, one for each of the post-strata groups. Each post-stratum group page will include the seven age-sex post-strata within the group, as well as a column for the group totals. If some post-strata are collapsed over age, then only the collapsed post-strata will be included in the tables. Twenty-five variables will be output for each post-stratum and post-stratum group total. The formulas to calculate the 25 items of interest are give below the table, along with the three methods of calculating the group totals.

Post-Stratum Definition	**PS (h) 10*g+1**	**...**	**PS (h) 10*g+7**	**Group Total [(10*g+1)+... +(10*g+7)]**	**Length**
Data-Defined Persons (DD)	(1)	...	(1)	(a)	I10
Insufficient Information (II)	(2)	...	(2)	(a)	I10
Total Persons (C*)	(3)	...	(3)	(a)	I10
Nonmover Sample Size	(4)	...	(4)	(a)	I8
Inmover Sample Size	(5)	...	(5)	(a)	I8
Outmover Sample Size	(6)	...	(6)	(a)	I8
Weighted Nonmovers (N_n)	(7)	...	(7)	(a)	F21.10
Weighted Inmovers (N_i)	(8)	...	(8)	(a)	F21.10
Weighted Outmovers (N_o)	(9)	...	(9)	(a)	F21.10
Weighted Nonmover Matches (M_n)	(10)	...	(10)	(a)	F21.10
Weighted Outmover Matches (M_n)	(11)	...	(11)	(a)	F21.10
Weighted P-Sample Persons (N_p)	(12)	...	(12)	(a)	F21.10
Weighted P-Sample Matches (M)	(13)	...	(13)	(a)	F21.10
E-Sample Size	(14)	...	(14)	(a)	I8
Correct Enumeration Sample Size	(15)	...	(15)	(a)	I8
Weighted E-Sample Persons (N_e)	(16)	...	(16)	(a)	F21.10
Weighted Correct Enumerations (CE)	(17)	...	(17)	(a)	F21.10
Dual System Estimate (DSE)	(18)	...	(18)	(a)	F21.10
Standard Error (SE(DSE))	(19)	...	(19)	(b)	F21.10
Coefficient of Variation (CV(DSE))	(20)	...	(20)	(c)	F12.10
Coverage Correction Factor (CCF)	(21)	...	(21)	(c)	F12.10
Standard Error (SE(CCF))	(22)	...	(22)	(c)	F12.10
Coefficient of Variation (CV(CCF))	(23)	...	(23)	(c)	F12.10

Net Undercount Percent (UC)	(24)	...	(24)	(c)	%13.8
Standard Error (SE(UC))	(25)	...	(25)	(c)	%12.8

(1) Data-Defined Persons = $C_h^{*} - II_h^{*}$

(2) Insufficient Information = II_h^{*}

(3) Total Persons = C_h^{*}

(4) Nonmover Sample Size = $\sum_{\substack{MOVERPER=1,\ RPROB^{(0)}>0\\ TESFINWT>0}} 1$

(5) Inmover Sample Size = $\sum_{MOVERPER=2,\ RSC\neq R} 1$

(6) Outmover Sample Size = $\sum_{\substack{MOVERPER=3,\ RPROB^{(0)}>0\\ TESFINWT>0}} 1$

(7) Weighted Nonmovers = $N_{n,h}^{(0)}$

(8) Weighted Inmovers = $N_{i,h}^{(0)}$

(9) Weighted Outmovers = $N_{o,h}^{(0)}$

(10) Weighted Nonmover Matches = $M_{n,h}^{(0)}$

(11) Weighted Outmover Matches = $M_{o,h}^{(0)}$

(12) Weighted P-Sample Persons = $\begin{cases} N_{n,h}^{(0)} + N_{i,h}^{(0)} & \text{if PES-C used} \\ N_{n,h}^{(0)} + N_{o,h}^{(0)} & \text{if PES-A used} \end{cases}$

(13) Weighted P-Sample Matches = $\begin{cases} M_{n,h}^{(0)} + \left(\dfrac{M_{o,h}^{(0)}}{N_{o,h}^{(0)}} \right) N_{i,h}^{(0)} & \text{if PES-C used} \\ M_{n,h}^{(0)} + M_{o,h}^{(0)} & \text{if PES-A used} \end{cases}$

(14) E-Sample Size = $\sum_{EWGHT>0,\ TESWGT>0} 1$

(15) Correct Enumeration Sample Size = $\sum_{CEPROBF^{(0)}>0} 1$

(16) Weighted E-Sample Persons = $N_{e,h}^{(0)}$

(17) Weighted Correct Enumerations = $CE_h^{(0)}$

(18) Dual System Estimate = $D\hat{S}E_h^{(0)}$

(19) Standard Error DSE = $\sqrt{Var(D\hat{S}E_h)}$

(20) Coefficient of Variation DSE = $\sqrt{Var(D\hat{S}E_h)} \ / \ D\hat{S}E_h^{(0)}$

(21) Coverage Correction Factor = $D\hat{S}E_h^{(0)} \ / \ C_h$

(22) Standard Error CCF = $\sqrt{Var(C\hat{C}F_h)}$

(23) Coefficient of Variation CCF = $\sqrt{Var(C\hat{C}F_h)} \ / \ C\hat{C}F_h^{(0)}$

(24) Net Undercount Percent $(\mathrm{UC_h}) = 100\% \times \dfrac{\mathrm{D\hat{S}E_h - C_h}}{\mathrm{D\hat{S}E_h}}$

(25) Standard Error UC = $(100\% - \mathrm{UC_h}) \times \mathrm{CV(D\hat{S}E_h)}$

For the post-stratum group totals,

(a) Item of Interest = $\sum_{h \in group}$ Item of Interest$_h$

(b) $\mathrm{DSE_{group}} = \sum_{h \in group} \sum_{h' \in group} \mathrm{Cov(D\hat{S}E_h, D\hat{S}E_{h'})}$

(c) Use formulas (19)-(25) on the group totals.

2. National Variance File

This file will only have one record, the national values.

National Variance File

Variable	Variable Description	Format
$\mathrm{DSE_{US}}$	DSE Estimate	F21.10
$\mathrm{Var(DSE_{US})}$	DSE Variance Estimate	F21.6
$\mathrm{SE(DSE_{US})}$	DSE Standard Error Estimate	F21.10
$\mathrm{CV(DSE_{US})}$	CV of the DSE	F12.10
$\mathrm{UC_{US}}$	Net Undercount Percent	%13.8
$\mathrm{SE(UC_{US})}$	UC Standard Error Estimate	%12.8

3. DSE & CCF Variance-Covariance Matrices (ASCII Format)

Output the DSE and CCF covariances, $\mathrm{Cov(D\hat{S}E_h, D\hat{S}E_{h'})}$ and $\mathrm{Cov(C\hat{C}F_h, C\hat{C}F_{h'})}$, to separate files with the following similar formats:

A.C.E. DSE Variance-Covariance Matrix

Variable	Variable Description	Format
h	Collapsed Post-Stratum Code	I3
h'	Collapsed Post-Stratum Code	I3
$\mathrm{Cov(D\hat{S}E_h, D\hat{S}E_{h'})}$	Covariance of DSE estimates for collapsed post-strata h and h'	F22.6

A.C.E. CCF Variance-Covariance Matrix

Variable	Variable Description	Format
h	Collapsed Post-Stratum Code	I3
h'	Collapsed Post-Stratum Code	I3
$Cov(\hat{CCF}_h, \hat{CCF}_{h'})$	Covariance of CCF estimates for collapsed post-strata h and h'	F13.10

For both files, (h,h') should run as: (1,1), (1,2), (1,3), ..., (K,K-1), (K,K). Thus, each file should have K^2 lines.

4. Replicate Weight File

With the definitions in section III.D.4 and the information in this file, it will be possible to compute the replicate weight for any replicate for any cluster. It will have 29,136 records. If the cluster was not in the final sample (11,303), then i, FSS, RW2, RW3, and RW4 should be blank.

Replicate Weight File

Variable	Variable Description	Format
j	Consecutive numbering of all 29,136 initial clusters	I5
i	Consecutive numbering of 11,303 clusters in final sample	I5
ISS	Initial Sampling Stratum Code (collapsed per III.D.3)	I3
FSS	Final Sampling Stratum Code (collapsed per III.D.3)	I7
SB	Small Block Subsampling Indicator	I1
RW2	Second Possible RW Value for Block Cluster i (from III.D.4)	F13.10
RW3	Third Possible RW Value for Block Cluster i (from III.D.4)	F13.10
RW4	Fourth Possible RW Value for Block Cluster i (from III.D.4)	F13.10

G. Variance Estimation Example

Below are examples of three steps of the process - replicate weight calculation, term estimation, and variance estimation.

1. Replicate Weight Example

Below is an example of a replicate weight file. There were initially 12 clusters, of which four were sampled out. There are two Initial Sampling Strata, with two Final Sampling Strata within each.

Cluster	j	i	ISS	FSS	N(ISS)	N(FSS)	R(FSS)
10001	1	1	1	11	7	3	2
10002	2	2	1	11	7	3	2
10003	3		1	11	7	3	2
10004	4	3	1	12	7	4	2
10005	5	4	1	12	7	4	2
10006	6		1	12	7	4	2
10007	7		1	12	7	4	2
10008	8	5	2	21	5	3	2
10009	9	6	2	21	5	3	2
10010	10		2	21	5	3	2
10011	11	7	2	22	5	2	2
10012	12	8	2	22	5	2	2

Note that there are thirteen replicates for the eight "surviving" clusters.

Cluster	j	i	ISS	FSS	N(ISS)	N(FSS)	R(FSS)	RW0	RW1	RW2	RW3	RW4	RW5	RW6	RW7
10001	1	1	1	11	7	3	2	1	0	(2/1)(2/3)(7/6)	(2/3)(7/6)	(7/6)	(7/6)	(7/6)	(7/6)
10002	2	2	1	11	7	3	2	1	(2/1)(2/3)(7/6)	0	(2/3)(7/6)	(7/6)	(7/6)	(7/6)	(7/6)
10004	4	3	1	12	7	4	2	1	(7/6)	(7/6)	(7/6)	0	(2/1)(3/4)(7/6)	(3/4)(7/6)	(3/4)(7/6)
10005	5	4	1	12	7	4	2	1	(7/6)	(7/6)	(7/6)	(2/1)(3/4)(7/6)	0	(3/4)(7/6)	(3/4)(7/6)
10008	8	5	2	21	5	3	2	1	1	1	1	1	1	1	1
10009	9	6	2	21	5	3	2	1	1	1	1	1	1	1	1
10011	11	7	2	22	5	2	2	1	1	1	1	1	1	1	1
10012	12	8	2	22	5	2	2	1	1	1	1	1	1	1	1

Cluster	j	i	ISS	FSS	N(ISS)	N(FSS)	R(FSS)	RW8	RW9	RW10	RW11	RW12
10001	1	1	1	11	7	3	2	1	1	1	1	1
10002	2	2	1	11	7	3	2	1	1	1	1	1
10004	4	3	1	12	7	4	2	1	1	1	1	1
10005	5	4	1	12	7	4	2	1	1	1	1	1
10008	8	5	2	21	5	3	2	0	(2/1)(2/3)(5/4)	(2/3)(5/4)	(5/4)	(5/4)
10009	9	6	2	21	5	3	2	(2/1)(2/3)(5/4)	0	(2/3)(5/4)	(5/4)	(5/4)
10011	11	7	2	22	5	2	2	(5/4)	(5/4)	(5/4)	0	(2/1)(½)(5/4)
10012	12	8	2	22	5	2	2	(5/4)	(5/4)	(5/4)	(2/1)(½)(5/4)	0

2. Term Estimation Example

Continuing the example above, for a specific post-stratum, h, and a specific term, k:

Cluster	j	i	TESELECT	$\sum w_{ipk} x_{ip}$	$\sum w_{ipk} y_{ip}$	$\sum w_{ipk} z_{ip}$
10001	1	1	U	400	100	0
10002	2	2	S	120	0	30
10004	4	3	N	300	0	0
10005	5	4	U	210	50	0
10008	8	5	S	640	0	110
10009	9	6	I	20	0	0
10011	11	7	I	90	0	0
10012	12	8	N	220	0	0

From the TESELECT codes, clusters 10001 and 10005 were selected with certainty, 10002 and 10008 were selected systematically, 10004 and 10012 were eligible for selection but were not selected, and 10009 and 10011 were not eligible for TES sampling. For replicates 0, 1, and 2:

$$\begin{aligned}
\mathrm{Term}_{k,h}^{(0)} &= (400 + 120 + 300 + 210 + 640 + 20 + 90 + 220) \\
&+ (100 + 50) \ + \ \frac{4}{2}(30 + 110) \\
&= 2430
\end{aligned}$$

$$\begin{aligned}
\mathrm{Term}_{k,h}^{(1)} &= (0 + \frac{2}{1}\frac{2}{3}\frac{7}{6}120 + \frac{7}{6}300 + \frac{7}{6}210 + 640 + 20 + 90 + 220) \\
&+ (0 + \frac{7}{6}50) \ + \ \frac{\frac{2}{1}\frac{2}{3}\frac{7}{6} + \frac{7}{6} + 1 + 1}{\frac{2}{1}\frac{2}{3}\frac{7}{6} + 1}(\frac{2}{1}\frac{2}{3}\frac{7}{6}30 + 110) \\
&= (1751\frac{2}{3}) \ + \ (58\frac{1}{3}) \ + \ \frac{4\frac{13}{18}}{2\frac{5}{9}}(156\frac{2}{3}) \\
&= 2099.4928
\end{aligned}$$

$$\begin{aligned}
\mathrm{Term}_{k,h}^{(2)} &= (\frac{2}{1}\frac{2}{3}\frac{7}{6}400 + 0 + \frac{7}{6}300 + \frac{7}{6}210 + 640 + 20 + 90 + 220) \\
&+ (\frac{2}{1}\frac{2}{3}\frac{7}{6}100 + \frac{7}{6}50) \ + \ \frac{0 + \frac{7}{6} + 1 + 1}{0 + 1}(0 + 110) \\
&= (2187\frac{2}{9}) \ + \ (213\frac{8}{9}) \ + \ \frac{3\frac{1}{6}}{1}(110) \\
&= 2749.4444
\end{aligned}$$

$$\begin{aligned}
\mathrm{Term}_{k,h}^{(3)} &= (\frac{2}{3}\frac{7}{6}400 + \frac{2}{3}\frac{7}{6}120 + \frac{7}{6}300 + \frac{7}{6}210 + 640 + 20 + 90 + 220) \\
&+ (\frac{2}{3}\frac{7}{6}100 + \frac{7}{6}50) \ + \ \frac{\frac{2}{3}\frac{7}{6} + \frac{7}{6} + 1 + 1}{\frac{2}{3}\frac{7}{6} + 1}(\frac{2}{3}\frac{7}{6}30 + 110) \\
&= (1969\frac{4}{9}) \ + \ (136\frac{1}{9}) \ + \ \frac{3\frac{17}{18}}{2\frac{5}{6}}(133\frac{1}{3}) \\
&= 2291.1765
\end{aligned}$$

3. Variance Estimation Example

Finishing the example, we have calculated the DSE for each of the 13 replicates, and we are now ready to estimate the variance of the DSE for this specific post-stratum, h.

			$(\mathrm{DSE}_h^{(j)}-$
j	N(ISS)	$\mathrm{DSE}_h^{(j)}$	$\mathrm{DSE}_h^{(0)})^2$
0		2500	
1	7	2440	3600
2	7	2570	4900
3	7	2520	400
4	7	2800	90000
5	7	2470	900
6	7	2490	100
7	7	2600	10000
8	5	2410	8100
9	5	2530	900
10	5	2550	2500

11	5	2490	100
12	5	2610	12100

$$\begin{aligned}\mathrm{Var}(DSE_h) &= \tfrac{6}{7}(3600 + 4900 + 400 + 90000 + 900 + 100 + 10000)\\ &+ \tfrac{4}{5}(8100 + 900 + 2500 + 100 + 12100)\\ &= 113160\end{aligned}$$

Let $C_h^* = 2450$. Then,

$$CCF_h = \frac{2500}{2450} = 1.0204$$

$$\mathrm{Var}(CCF_h) = \frac{\mathrm{Var}(DSE_h)}{2450^2} = \frac{113160}{2450^2} = 0.0189$$

G. Testing and Verification

1. Testing

a. First Phase: Correctness

This testing phase will be on a small dataset, taken from either the 1990 Post-Enumeration Survey or the Census 2000 Dress Rehearsal. This will ensure that the program is operating correctly.

b. Second Phase: Scale

This testing phase will be on a dataset comparable in size to what we expect the production system will have to handle. This testing will ensure that there are no unexpected issues due to the very large size of the input files (e.g. the program takes 3 days to run). P- and E-sample missing data files based on the Census 2000 Dress Rehearsal have been produced for this purpose.

2. Verification

Verification of the variance estimation output will be accomplished by comparing the "production" outputs to those from an independently programmed verification program created by Robert Sands and Roger Shores. The A.C.E. Post-Stratum Direct Variance file can also be compared to Collapsed Post-Stratum-Level file, an output from the DSE estimation operation. Verification will take place during the testing phase, as well as during official production.